File Number: 25-13

Joint
Computer-aided Acquisition
and Logistic Support (JCALS)
CALS Technology Center (CTC)

Requirements Paper

Approach for Sending JEDMICS/CADA Rejected Images to Pending Storage for Manual Review

CONTRACT NO.: DAAB07-93-D-T001
TASK NO.: 96-23

15 November 1996

Prepared for:

Department of the Army
PM JCALS

Prepared By:

ACCURATE Information Systems, Inc.
Meridian Center 1
2 Industrial Way West
Eatontown, New Jersey 07724

Joint Engineering Data Management Information And Control System/ Computer-Assisted Data Acceptance (JEDMICS/CADA)

Requirments Paper

Approach for Sending JEDMICS/CADA Rejected Images to Pending Storage for Manual Review

15 November 1996

FILE NUMBER: 25-13

APPROACH FOR SENDING JEDMICS/CADA REJECTED IMAGES TO PENDING STORAGE FOR MANUAL REVIEW

1. INTRODUCTION

The Joint Engineering Data Management Information Control Systems/Computer Assisted Data Acceptance (JEDMICS/CADA) system performs automated Quality Assurance (QA) on raster engineering drawings. This system operates in a non-intrusive manner with the host JEDMICS repository by using the JEDMICS Application Program Interface (API) to extract image files from either Permanent or Pending storage.

The automated QA involves decisions relative to Image Quality, Image Validation, and Index Verification. After these decisions are completed, an evaluation status report, which contains the results of the evaluation, may be printed. This report can contain the entire set of images, or just those which were marked as "Rejects." A rejected image is one that fails to pass at least one of the criteria of the automated evaluation. The report contains the JEDMICS index information necessary to query each image. This makes it possible for QA operators to select the images on the report and view them at a JEDMICS workstation later.

If the images had originated from Pending, the evaluated batch can be output as a new batch to Pending. In this case, the automated evaluation results would be mapped to the appropriate JEDMICS QA flags. The new batch can then be viewed on a JEDMICS workstation.

This paper will address using JEDMICS/CADA to query Permanent, perform evaluation, and output the evaluation results as a new batch to Pending.

2. JEDMICS/CADA FUNCTIONAL CONFIGURATION

The functional configuration of the current JEDMICS/CADA system is shown in Figure 1. The system can initiate a query to either JEDMICS Pending or Permanent via the JEDMICS API (version 2.5). As shown, read only access of Permanent is allowed while read/write access can occur to Pending. The capabilities for automated image and index processing are shown, as well as operator visual QA and output to a printer or uploaded as a new batch to Pending.

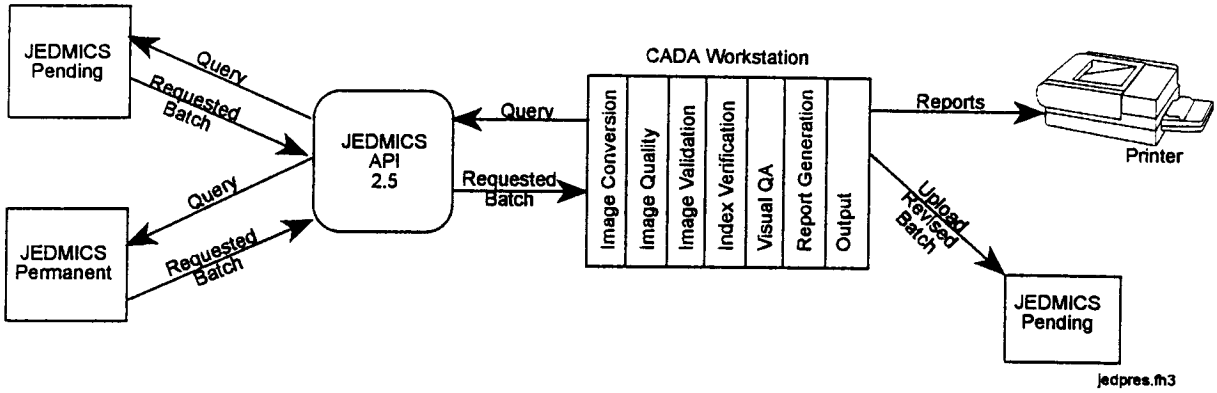

Figure 1. JEDMICS/CADA Functional Configuration

3. APPROACH

The approach for performing automated QA against the JEDMICS Permanent data base is comprised of the following steps

1. **Identification of Query**: JEDMICS/CADA uses the JEDMICS API for the identification and extraction of images. The API has a limitation with respect to the number of hits a query can return. It has been recommended that this limit be below 1,000 to ensure that JEDMICS system performance is not compromised. For this reason, all Permanent queries from JEDMICS/CADA must be manually coordinated by the operator. In this case, where all of Permanent is attempting to be processed, it is recommended that it be performed on a platter by platter basis.

2. **Query against Permanent**: The operator will submit the determined query against Permanent. JEDMICS/CADA will then process the query results by extracting the images in sub-batches. The sub-batch size, recommended to be 100, can be changed by the operator.

3. **Perform QA**: JEDMICS/CADA will now perform automated QA on each image retrieved. The QA will consist of the following.

format check -	This checks the C4 format.
image quality-	This checks the legibility of the image from an image quality standpoint. It is recommended that the quality checks be performed for legacy data. This is a less strict check.
image validation-	This identifies the image borders, checks the image dimensions, and determines the image size on the border to border dimensions.
index verification-	This verifies that the key ID data contained in the JEDMICS index matches the same index found on the image.

4. **Save "rejected images"**: The JEDMICS API does not allow for the update of index information or the addition of images to an existing Pending batch. This means that all images must be loaded into the batch at the time of batch creation. For this reason, JEDMICS/CADA must save all rejected images locally until enough images are obtained to constitute creating a new batch on Pending.

5. **Map to QA flags**: Once a batch is evaluated, the JEDMICS/CADA decisions will be mapped to the appropriate QA flags. A fair amount of detail is lost during this mapping. JEDMICS/CADA provides details regarding where the image quality failed and specific Hollerith fields where ID verification failed.

   For this reason it is recommended that the evaluation status report, generated by JEDMICS/CADA be printed and referenced during manual QA.

6. **Output to Pending**: A new batch will be generated on Pending and the identified rejected images will be inserted into the batch. A Pending Status report will also be generated which will inform the QA operators which new batch was created and which images were placed in the batch. This is recommended for accountability purposes.

4. THE FEASIBILITY AND PRACTICALITY OF THIS IMPLEMENTATION

4.1 Feasibility

The ability to query images from Permanent, perform automated evaluation, and output the evaluation results to Pending is feasible. Currently, JEDMICS/CADA has proven the feasibility by having the basic components necessary for this implementation. These components are the ability to read from both data bases and to create and populate new batches in Pending.

If this approach were taken, the following modifications to the existing JEDMICS/CADA system would be necessary.

1. **Output only rejected images**- JEDMICS/CADA would now be required for outputting only those images that failed during automated evaluation. Due to API constraints, JEDMICS/CADA does not have the ability to modify or add to an existing Pending batch. This means that all images must be output to a batch when it is created. This will cause JEDMICS/CADA to store all rejected image files and index information locally until an appropriate batch is present. At that time, the new Pending batch will be created and populated.

2. **Output images from Permanent to Pending** - JEDMICS/CADA will now need to allow for the creation of Pending batches from a Permanent source. The current version of JEDMICS/CADA does not allow for this. This modification can be made.

4.2 Practicality

The approach, as described in this paper, poses some problems. This section addresses the present concerns.

4.2.1 Magnitude of Manual Inspection

If the full set of JEDMICS/CADA capabilities is exercised (i.e., C4 format validation, image quality and index verification) the number of expected rejected images could be high. Limited testing using the CECOM JEDMICS system, with full JEDMICS/CADA processing employed, revealed that the average percentage of rejects could approach 60 to 70 percent. Assuming a repository of 1,000,000 drawings, this would result in seven-hundred thousand images that would require manual inspection. This puts a tremendous strain on JEDMICS personnel.

4.2.2 Number of Batches loaded to Pending

A large number of Pending batches will be needed to accommodate the estimated number of images that will require manual inspection. This will cause a strain on the current processing Pending performs for loading new data. It is also important to remember that the Pending data base was not intended to be used in this manner. New batches do not reside for long in the Pending data base. If this approach is followed, there is a strong potential for the batches created from Permanent to be resident in Pending for long periods of time. This is because these batches must undergo a manual inspection, and if deemed unacceptable, must be deleted from Permanent and input into the system again. Needless to say, in some cases this may take a great deal of time, and greatly impact JEDMICS' normal repository business process.

Based on this analysis, and the limited testing performed, it is not practical to retrieve batch data from Permanent, perform full CADA processing, and input the rejected images as new batches into Pending. The JEDMICS/CADA system can be more effectively used for format and Image QA of the Permanent data.

4.3 Recommendation

Although this approach is feasible in implementation, it is not the most practical. It would cause a great deal of processing on the Pending data base and require an enormous amount of human intervention to release the JEDMICS/CADA created batches. The expected volume of images that would have to be dealt with by the operators makes this an impractical option for implementation.

Section 5 discusses a more practical way of ensuring good quality of the JEDMICS repository.

5. ALTERNATIVE APPROACH

The goal for performing QA is to obtain good quality data. This goal can be met in ways other than performing QA on all images stored within the JEDMICS repository. A more practical use of the JEDMICS/CADA system would be to use strict QA checks against all new data being entered into the

repository and against all data being output (bid sets or other). In addition, image checks to identify image problems that may have occurred during migration can be run against the entire repository.

This advantages in using the JEDMICS/CADA system in this manner are discussed in the Section 5.1 through 5.3.

5.1 Detailed QA on Input

The JEDMICS/CADA system should be used to provide QA on all new batches received through Pending. The identified image quality problems, format problems, and index problems will be provided as new batches within Pending that can be qualified by the QA operators. This will greatly reduce manual QA by the operators. Operators may perform selective manual QA based on automated reject reasons, contained in the QA flags, or simply perform a random sampling. The operators still have the only authority for releasing batches to Permanent.

5.2 Selective QA of Permanent

It is recommended that a more focused approach be made in the QA of the large amount of engineering drawing images within Permanent. The ability to detect compression and image corruption problems exists within JEDMICS/CADA. By performing a selective QA of all images within Permanent, which identifies only severe image problems (the image is all black or all white) and format problems, the amount of manual QA required to handle the rejected images is reduced.

JEDMICS/CADA provides a wide variety of capabilities for reporting QA problems that cannot be mapped to JEDMICS QA flags. For this reason, it is recommended that the evaluation status report which is generated by JEDMICS/CADA be used by the QA operator when viewing the problem images. This report should be used in conjunction with viewing the images on the JEDMICS QA workstation.

If there were a means within the JEDMICS index of flagging an image as having undergone automated evaluation, then the full QA could be run. If the image passed, it could be flagged accordingly, which would save the QA on output. Then, only those images that had failed the selective QA would be output to Pending. This index flag, however, does not currently exist in the index information.

5.3 Detailed QA on Output

By providing a selective QA of Permanent and detailed QA on input and output, good data quality will be ensured. This approach is practical because images will now be checked on demand. It is estimated that between 60 and 80 percent of the existing images within a JEDMICS repository will never be accessed. Performing QA on all output sets ensures good quality before final output to the media devices and reduces the amount of manual inspection of rejected images.

This approach for performing detailed QA on output is discussed in a separate report titled *JEDMICS/CADA Quality Assurance of Output from JEDMICS.*

6 CONCLUSIONS AND RECOMMENDATIONS

The conclusion reached is that the original idea of performing automated QA on JEDMICS Permanent is feasible. The recommendation, however, is to use the capabilities of JEDMICS/CADA in a more focused manner. This approach is to perform all JEDMICS/CADA processing on both the input and output sides of JEDMICS and perform only format and corruption checks against the full Permanent.

JEDMICS/CADA also provides a greater level of detail in reporting QA problems than the JEDMICS QA flags allow. Because of this some detail is lost when the JEDMICS/CADA decisions are mapped to the JEDMICS QA flags. For this reason it is recommended that when QA is performed and the rejected output is returned to Pending, the evaluation status report should also be consulted.